

NAVAL POSTGRADUATE SCHOOL

Monterey, California



THESIS

**CORRELATION ANALYSIS OF FLEET INFORMATION
WARFARE CENTER NETWORK INCIDENTS**

by

Patrick W. Ginn

June 2001

Thesis Advisor:

Associate Advisor:

Raymond Buettner

Dan Boger

Approved for public release; distribution is unlimited

20011116 160

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED	
	June 2001	Master's Thesis	
4. TITLE AND SUBTITLE: Title (Mix case letters)		5. FUNDING NUMBERS	
Correlation Analysis of Fleet Information Warfare Center Network Incidents			
6. AUTHOR(S) Patrick W. Ginn			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)		8. PERFORMING ORGANIZATION REPORT NUMBER	
Naval Postgraduate School Monterey, CA 93943-5000			
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
N/A			
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT		12b. DISTRIBUTION CODE	
Approved for public release; distribution is unlimited.		A	
13. ABSTRACT (maximum 200 words)			
<p>The Navy's Intrusion Detection process is currently reactive in nature. It is designed and programmed to detect and provide alerts to the Fleet Information Warfare Center (FIWC) of suspicious network activity while it is in progress, as well as to record/store data for future reference. However, the majority of activity taking place within and across Naval networks is legitimate and not an unauthorized activity. To allow for efficient access and utilization of the information systems sharing the network the Intrusion Detection Systems must be set at a level that filters out activity deemed as normal or non-hostile, while still providing an appropriate level of security. With this filtering in place an IDS system will not register all suspicious activity, and may not detect mild and seemingly harmless activity. When increasing security, limits must be imposed upon access. This thesis examines FIWC network incident data from 1999 to see if a correlation can be drawn between United States visibility in the foreign media during 1999 and the occurrence of suspicious network incidents. A positive correlation may provide advance-warning indicators that could lead to the development of a procedure for increasing security posture based on the current environment. These indicators would provide a more proactive method of defense, significantly reduce potential damage caused by hostile network incidents and provide for more efficient network activity.</p>			
14. SUBJECT TERMS		15. NUMBER OF PAGES	
Fleet Information Warfare Center, Intrusion Detection, Network Incidents		74	
16. PRICE CODE			
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

**CORRELATION ANALYSIS OF FLEET INFORMATION
WARFARE CENTER NETWORK INCIDENTS**

Patrick W. Ginn
Captain, United States Army
B.B.A., North Georgia College, 1991

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN
INFORMATION TECHNOLOGY MANAGEMENT**

from the

NAVAL POSTGRADUATE SCHOOL
June 2001

Author:


Patrick W. Ginn

Approved by:


Raymond Buettner, Thesis Advisor


Dan Boger, Associate Advisor.


Dan Boger, Chairman
Information Systems Academic Group

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Navy's Intrusion Detection process is currently reactive in nature. It is designed and programmed to detect and provide alerts to the Fleet Information Warfare Center (FIWC) of suspicious network activity while it is in progress, as well as to record/store data for future reference. However, the majority of activity taking place within and across Naval networks is legitimate and not an unauthorized activity. To allow for efficient access and utilization of the information systems sharing the network Intrusion Detection Systems must be set at a level that filters out activity deemed as normal or non-hostile, while still providing an appropriate level of security. With this filtering in place an IDS system will not register all suspicious activity, and may not detect mild and seemingly harmless activity. When increasing security, limits must be imposed upon access. This thesis examines FIWC network incident data from 1999 to see if a correlation can be drawn between United States visibility in the foreign media during 1999 and the occurrence of suspicious network incidents. A positive correlation may provide advance-warning indicators that could lead to the development of a procedure for increasing security posture based on the current environment. These indicators would provide a more proactive method of defense, significantly reduce potential damage caused by hostile network incidents and provide for more efficient network activity.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	BACKGROUND	1
B.	STATEMENT OF PROBLEM.....	3
1.	Research Questions.....	4
2.	Scope and Assumptions	4
C.	RESEARCH OBJECTIVES.....	5
II.	FLEET INFORMATION WARFARE CENTER	7
A.	OVERVIEW.....	7
1.	History.....	7
2.	Mission and Structure	7
B.	NETWORK SECURITY.....	8
1.	Naval Computer Incident Response Team.....	8
2.	Intrusion Detection	10
III.	PACIFIC NORTHWEST NATIONAL LABORATORIES.....	13
A.	OVERVIEW.....	13
1.	History.....	13
2.	Mission	13
B.	VISUAL INFORMATION ANALYSIS	14
1.	Overview	14
2.	Spacial Paradigm for Information Retrieval and Exploration	14
IV.	DATA	19
A.	DESCRIPTION AND SELECTION.....	19
1.	Naval Network Incidents	19
2.	Foreign Broadcast Information Service.....	20
B.	DATA AGGREGATION	21
1.	Naval Network Incidents	21
2.	Media Visibility	23
V.	METHODOLOGY	27
A.	OVERVIEW.....	27
B.	HYPOTHESIS TESTING.....	28
C.	AUTOCORRELATION.....	30
D.	CORRELATION	32
E.	PROCESS	35
F.	SPIRE ANALYSIS TOOLS.....	36
VI.	ANALYSIS	39
VII.	CONCLUSIONS AND RECOMMENDATIONS.....	47
A.	SUMMARY	47
B.	CONCLUSIONS	48
C.	RECOMMENDATIONS FOR FURTHER RESEARCH	48

LIST OF REFERENCES	51
INITIAL DISTRIBUTION LIST	53

LIST OF FIGURES

Figure 2.1	U.S. Navy Organizational Structure	8
Figure 2.2	NAVCIRT Organization Structure	9
Figure 3.1	Galaxies Visualization	16
Figure 3.2	ThemeView Visualization	16
Figure 4.1	Network Incidents	22
Figure 4.2	Document Visibility	24
Figure 4.3	FBIS Time Slices	25
Figure 6.1	FBIS Documents by Quarter	41
Figure 6.2	FBIS Documents by Quarter	41
Figure 6.3	FBIS Documents by Quarter	42
Figure 6.4	FBIS Documents by Quarter	42
Figure 6.5	Incident & Visibility Graph	45

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 4.1	Incident Category Description	19
Table 4.2	Summary Statistics.....	22
Table 4.3	Summary Statistics.....	24

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS

CERT	Computer Emergency Response Team
CINCLANTFLT	Commander in Chief Atlantic Fleet
CINCPACFLT	Commander in Chief Pacific Fleet
CINCUSNAVEUR	Commander and Chief US Navy Europe
CND	Computer Network Defense
CNO	Chief Naval Operations
COMUSNAVCENT	Commander US Navy Central
COMNAVSECGRU COM	Commander Navy Security Group Command
DOD	Department of Defense
DON	Department of Navy
IA	Information Assurance
IDS	Intrusion Detection System
IO	Information Operations
IW	Information Warfare
FBIS	Foreign Broadcast Information Service
FIWC	Fleet Information Warfare Center
JTFCND	Joint Task Force Computer Network Defense
NAVCIRT	Naval Computer Incident Response Team
PNL	Pacific Northwest Laboratory
SPAWAR	Space and Naval Warfare Systems Command
SPIRE	Spacial Paradigm for Information Retrieval and Exploration
VIA	Visual Information Analysis

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to express my thanks to the Navy's Fleet Information Warfare Center for providing me with the funding to travel and conduct research, and to Pacific Northwest National Laboratory for providing me with the software analysis tool. I would also like to express my thanks to LT Ray Buettner for his guidance and patience, and to Dr. Dan Boger for his valued insight and input as a second reader.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

Information, information processing, and communications networks are at the core of every military activity. The military forces of the United States rely upon real-time information to acquire and maintain an information dominance and superiority. It is information superiority that allows the capability to collect, process, and disseminate an uninterrupted flow of information while exploiting or denying an adversary's ability to do the same (JV 2020). However, having information superiority does not equate to having perfect or accurate information. Collected data will only be as good as the systems and processes that are in place to ensure its reliability and accuracy. Only then can coherent information be gathered to provide a competitive advantage by translation into superior knowledge and decisions.

Information Operations (IO) are actions taken to affect an adversary's information and information systems while defending one's own information and information systems. It is segmented into Offensive and Defensive arenas, with Defensive IO focused on integrating and coordinating policies and procedures, operations, personnel, and technology to protect and defend information and information systems (Joint Pub 3-13). Defensive IO is conducted to ensure timely, accurate, and relevant information access while denying exploitation of friendly information and information systems. Information Assurance (IA) is the function of Defensive IO that seeks to protect and defend information and information systems by ensuring their availability, integrity, identification and authentication, confidentiality, and non-repudiation. IA employs

technologies and processes such as multilevel security, access controls, secure network servers, and intrusion detection software.

One of the ways the Navy conducts Information Assurance is through the Fleet Information Warfare Center (FIWC), which provides network Intrusion Detection System (IDS) coverage for the Navy. Utilizing the Cisco Secure IDS (formerly Net Ranger), as well as email and live phone reports from network administrators at Naval Network Operation Centers (NOCs) and Naval installations, Navy network incident data is transmitted to and collected by the Naval Computer Incident Response Team (NAVCIRT) operations center located at Norfolk, Virginia. There it is analyzed and verified by network technicians, then compiled to formulate a current situation picture and develop a historical database. All verified incidents are recorded, processed and passed on to the Joint Task Force for Computer Network Defense (JTF-CND). This manual analysis is utilized to determine significant events and to help reduce false alerts caused by legitimate network activity and random noise.

Pacific Northwest National Laboratory, an activity of the U.S. Department of Energy, has developed automated tools for data-mining vast amounts of textual information. These software tools are capable of searching for key words and phrases within data sources, and then presenting the results in a graphic theme schematic. This format allows for the viewing of seemingly unrelated textual data, which is clustered and grouped together based on common themes, key words or phrases found within the data and presented in a three dimensional display. Currently, the Land Information Warfare Activity (LIWA) for the U.S. Army and the Joint Information Operations Center (JIOC) utilize these tools.

The Foreign Broadcast Information Service (FBIS) Reports produced by the Central Intelligence Agency provide for an extensive amount of data concerning the foreign media. FBIS offers an extensive, in-depth collection of translations and transcriptions of open source information monitored worldwide on such diverse topics as military affairs, politics, the environment, societal issues, economics, and science and technology. The information is obtained monitoring radio, television, press, periodicals, books and other sources of unrestricted information such as databases and gray literature.

B. STATEMENT OF PROBLEM

Intrusion detection is a process that is reactive in nature. It is designed and to detect and provide alerts to automated monitoring systems and network administrators of suspicious network activity while it is in progress, or after it has occurred. Being able to determine why something happened would make it possible to determine when it might happen again. The hypothesis is that the visibility of the United States within the foreign media is directly correlated to the frequency of network incidents seen on Navy networks. The null hypothesis is that the two are not related.

Time series analysis is based on the premise that successive observations are usually not independent of one another. When successive observations are dependent, future values may be predicted from past observations. It is important to note that in most cases the future is only partially determined by previous events. Exact predictions are impossible, however, future values will have probability distributions based on past occurrences. This thesis examines FIWC network incident data from 1999 to see if a correlation can be drawn between United States visibility in the foreign media during 1999 and the occurrence of detected network incidents. A positive correlation may

provide advance-warning indicators that could lead to the development of a procedure for increasing security posture based on the current environment. These indicators would provide a more proactive method of defense, significantly reduce potential damage caused by hostile network incidents and provide for more efficient network activity.

1. Research Questions

- a. Can a methodology be developed to provide for a more proactive process of network defense?
- b. Is there a correlation between network incidents on Navy networks and visibility of the United States in the foreign media?
- c. Is United States visibility within the foreign media an event determinant?

2. Scope and Assumptions

This thesis will focus on FIWC data from recorded network incidents and foreign media events during 1999. The year 1999 will provide a complete history of recorded Naval network incident data from all fleet and shore installations; and the foreign media data from every country processed by the Foreign Broadcast Information Service. This will not be a technically focused discussion of what type of security measures to implement or how a network should be managed, but an analysis of incidents and foreign media events that transpired at approximately the same time to determine possible correlation. It is impossible to determine the speed of information flow. The Internet has provided near real-time access to news and events with transmission and response time being as fast as the click of a button. A logical assumption is that the majority of incidents are by trained adversaries with a formulated plan. With this in mind the correlation

planning figure used for lead-time between incident and media visibility is approximately 48 hours.

The advent and growing expansion of the World Wide Web and the Internet has made geographical boundaries a thing of the past. For example, a terrorist organization with religious or political ties to a country in the Middle East could actually be located within the continental United States or anywhere in the world and still conduct cyber attacks. During the investigation and verification of network incidents NAVCIRT does determine the actual originating source location of the incident. However, this specific location information makes the incident itself classified, so for this thesis actual physical location of attack is immaterial.

United States visibility in the foreign media will be used to correlate incidents or attacks. As stated, it is possible to determine the actual physical location of origin, but that alone is not enough to prove association to a foreign country. The assumption here is that foreign media is similar to domestic media, and that visibility in the press is a valid indicator regardless of media source. The same analysis could be done using domestic media events.

C. RESEARCH OBJECTIVES

Analyze network incidents and “visible” media events to establish whether a correlation exists between the two. If a correlation can be determined between United States visibility and network incidents, then a procedure for increasing security posture based on the current environment may be developed. This would help to significantly reduce potential damage caused by hostile network incidents. Specifically, the Fleet

Information Warfare Center is looking into the area of proactive network defense to better protect the Navy's information assets. A proactive approach to network defense could be utilized service wide to potentially provide regional warnings for increased security posture.

II. FLEET INFORMATION WARFARE CENTER

A. OVERVIEW

1. History

FIWC assumed the role of Navy Computer Incident Response Team (NAVCIRT) from SPAWAR on 1 October 1995. FIWC Headquarters and its subordinate detachments provide Information Warfare operational support to naval forces worldwide. In-service engineering Information Security (INFOSEC) support to include: hardware/software installations, firewall design and installations, product reviews, technical assistance, and acquisition of INFOSEC-related products is still conducted by SPAWAR Systems Center Charleston. In its brief history, the FIWC team has developed Information Warfare (IW) doctrine and tactics, provided highly skilled IW operational teams to deploying naval units, and pioneered defensive computer network support throughout the Navy and Marine Corps.

2. Mission and Structure

The Fleet Information Warfare Center's Naval Computer Incident Response Team (NAVCIRT), focuses on ensuring Navy and Marine Corps computer systems are capable of providing complete and unaltered information while withstanding malicious (internal and external) disruptions and attacks. It's mission is to provide continuously improving Information Warfare (IW) support, which encompasses and integrates all IW disciplines, to deploying and operating Naval forces worldwide as tasked by CINCLANTFLT, CINCPACFLT, CINCUSNAVEUR, and COMUSNAVCENT. FIWC is under the operational control of CINCLANTFLT (N39) (see Figure 2.1).

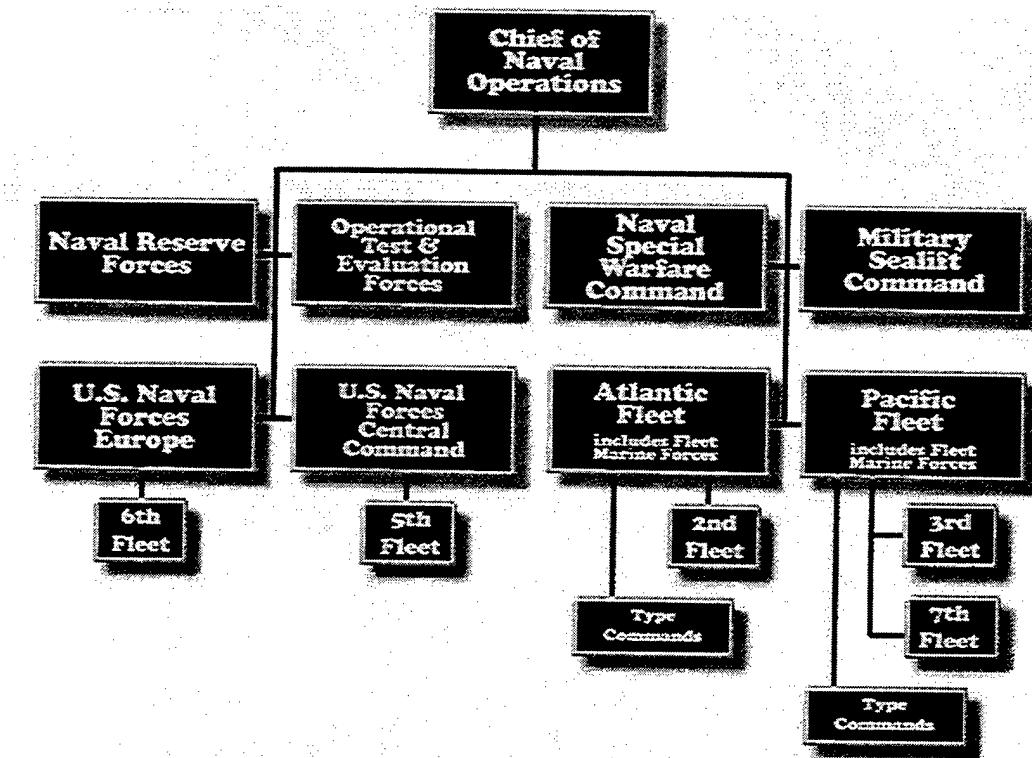


Figure 2.1 U.S. Navy Organizational Structure (From: Reference 10)

The NAVCIRT office was chartered under CNO ALCOM 45/95 to be the Navy and Marine Corps' single point of contact for reporting and handling computer security incidents and vulnerabilities. As the Naval Center of Excellence of Information Warfare (IW), FIWC is the preeminent command engaged in expanding the tactical commanders advantage in information exploitation, protection, and attack.

B. NETWORK SECURITY

1. Naval Computer Incident Response Team

The Naval Computer Incident Response Team (NAVCIRT), operated by the Fleet Information Warfare Center, provides computer security incident support to Department of Navy (DON) activities. As designated by OPNAVINST 2201.2, Navy and Marine Corps Computer Network Incident Response, NAVCIRT serves as the department of the

Navy (DON) focal point for the Department of Defense (DOD) exchange on computer technical vulnerabilities, viruses, and "hacker" incidents. NAVCIRT is also a member of the Forum of Incident Response and Security Teams (FIRST), which is comprised of international government agencies and commercial firms. NAVCIRT also provides computer security advisories which contain current vulnerability announcements on various systems, antiviral software information, and information on hacker incidents. There are four NOCs that are in direct support to the Navy fleets, broken down into four theater regions: Atlantic, Pacific, European, and Indian. The resource sponsor for FIWC is the Director of Space and Electronic Warfare (CNO N6). FIWC is authorized direct liaison with fleet commanders in chief, numbered fleets, type commanders, COMNAVSECGRU COM, carrier and cruiser/destroyer group commanders, navy laboratories, Commander Operational Test and Evaluation Force, and other agencies.

The Operations (N3) department (see Figure 2.2) is organized under the Directorate of

NAVCIRT OPERATIONS DEPARTMENT

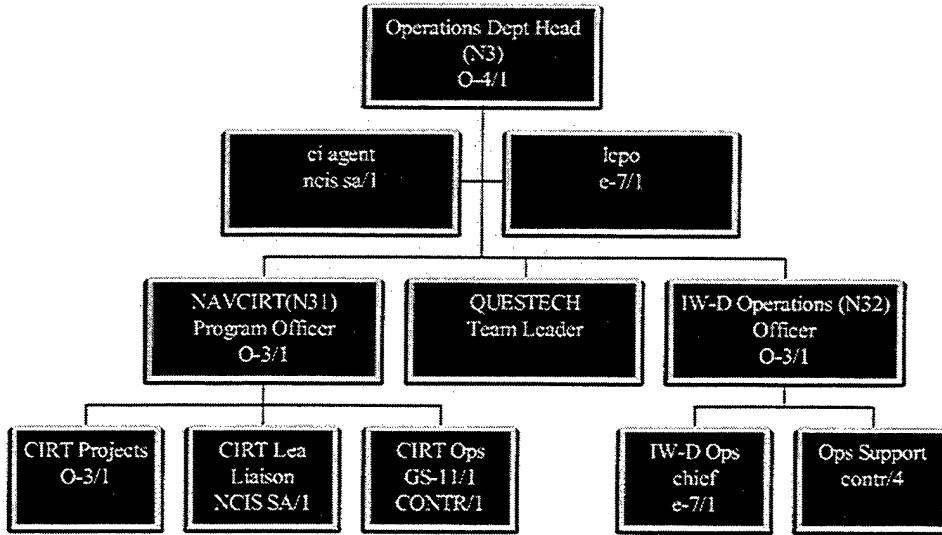


Figure 2.2 NAVCIRT Organization Structure

Operations, Plans, and Requirements. The department is divided into two divisions: NAVCIRT (N31) and Information Warfare-Defensive (IW-D) Operations (N32). NAVCIRT personnel coordinate actions involving security incidents and vulnerabilities, maintain the incident database, track all open incidents, provide technical support to customers, provide advisories on vulnerabilities to the fleet, and provide other technical support as required.

2. Intrusion Detection

Computer systems bring together a multitude of vulnerabilities. Hardware vulnerabilities are shared among the computer, the communication facilities, and the remote units and consoles, while software vulnerabilities are at all levels of the machine operating system and supporting software. There are three major categories of vulnerabilities: accidental disclosures, deliberate penetrations, and physical attack. And although an organization must be concerned with all three, it is deliberate penetration that provides the need for intrusion detection.

FIWC's Computer Network Defense Division uses various Intrusion Detection Systems (IDS) and processes to monitor Navy networks worldwide. Although FIWC is the center for network defense for the Navy, they are supported by four Naval Network Operation Centers (NOCs) that provide direct support to the Fleet. These NOCs provide regional network support based on a unit's location. Support to units will change as the unit transitions between regions. This transition between NOCs is transparent to the user.

An IDS will provide real-time intrusion alerts, designed to detect and report unauthorized activity throughout a network from a centralized environment. The Cisco Secure Intrusion Detection System (CSIDS), formerly known as NetRanger, works and

records network activity on a continuous basis and can operate in both Internet and intranet environments to protect an organizations' entire network. The IDS includes two components: Sensor and Director. The sensors, are high-speed network devices that analyze the content and context of individual packets to determine if traffic is authorized. While the Director terminals are located at the four Network Operating Centers, the Net Ranger sensors are distributed throughout the fleet and attached to all gateways between the Navy's Intranet and the Internet. If unauthorized activity is detected the sensor forwards alarms to the director management console. It can monitor almost any type of TCP/IP network, ranging from Internet connections, LAN segments, or dial-in modem pools. The Director is a high-performance software-based management system that centrally monitors the activity of multiple Sensors located on local or remote network segments. The ability to monitor many sensors and retain consistent quality of security allows network administrators to pinpoint the location and type of an attack and respond.

THIS PAGE INTENTIONALLY LEFT BLANK

III. PACIFIC NORTHWEST NATIONAL LABORATORIES

A. OVERVIEW

1. History

Pacific Northwest National Laboratory (PNL) began in 1965, and was contracted to perform research and development for the Hanford Nuclear Site in southeastern Washington State. The Laboratory's first projects were based on the needs of the Hanford Site and included protecting the environment, fabricating reactor fuel, and designing reactors. These projects, staff expertise in diverse fields, and national needs led to outstanding research and development in several key areas: environment, health, energy, computer science, and security. In the early 1990s, Pacific Northwest Laboratory became one of the multiprogram laboratories in the U.S. Department of Energy's national laboratory system. These labs augment America's existing academic and industrial research infrastructure. Research at the laboratories addresses national needs in such areas as environment, national security, health, manufacturing, high-performance computing, advanced materials, and other areas. Today, Pacific Northwest combines capabilities from a multitude of scientific and engineering disciplines to aid in the development of software tools to speed the discovery process. Their goal is to develop and enhance information systems and technologies by providing; data and knowledge engineering, real-time data acquisition and analysis and synthesis analysis and visualization of information.

2. Mission

Pacific Northwest National Laboratory's core mission is to deliver environmental science and technology in the service of the nation and humanity. Through basic research

they create fundamental knowledge of natural, engineered, and social systems that is the basis for both effective environmental technology and sound public policy. They also apply their capabilities to meet selected national security, energy, and human health needs; strengthen the U.S. economy; and support the education of future scientists and engineers. This mission supports the U.S. Department of Energy's Strategic Laboratory Mission Plan, where Pacific Northwest is designated as a principal laboratory in the environmental quality mission, a major contributing laboratory in both science and technology and energy, and a participating laboratory in national security.

B. VISUAL INFORMATION ANALYSIS

1. Overview

Visual Information Analysis (VIA) is the use of advanced computer graphics technologies to support the interactive visualization and exploration of large collections of complex information. More than just information visualization, Visual Information Analysis is the process of both examining and interacting with dynamic visual representations of information. The key components of any VIA system are 1) the information model the system uses to internally represent the contents of, and interrelationships among, the elements of an information collection, and 2) the visualization tools and user interface used to graphically represent (and support interaction with) the information model. An effective VIA system will incorporate both a rich, flexible data model, and a comprehensive set of powerful visual information analysis tools.

2. Spacial Paradigm for Information Retrieval and Exploration

Spacial Paradigm for Information Retrieval and Exploration (SPIRE) is a Visual Information Analysis tool designed to help with the identification of trends, patterns or

unexpected occurrences of themes or topics. Specifically, SPIRE provides the user with a macro-level view of thematic changes in a collection of documents. Users are able to interact with the visualization to explore the information to investigate apparent or unusual patterns, and other features of interest. The VIA represents each document or group of documents with an icon for a particular time period. Using SPIRE's visualization feature, a user may view a grouping of documents that span a specific period of time, or common theme. This type of visualization is called document-centric.

SPIRE is a new type of software that allows you to explore complex relationships among documents in a textual dataset. SPIRE is unique because it provides a visual interface to your data. Not only can you explore relationships by simply examining visualizations, but also you can interactively sift through documents to refine searches and explore temporal relationships in your data. SPIRE uses two technologies to perform this task: Galaxies and ThemeView visualizations.

The Galaxies Visualization (see Figure 3.1) is a scatter plot visualization in which documents are distributed across a two dimensional space. In the Galaxies visualization, documents are positioned so that they are located close to other highly related documents. By examining a Galaxies visualization, you can rapidly find clusters of related documents and explore inter-document relationships. The ThemeView Visualization (see Figure 3.2) is a content-based visualization derived from the Galaxies Visualization. In a ThemeView visualization, the dataset is represented as a landscape in which peaks indicate areas of high thematic content. The Galaxies visualization is a revolutionary approach to understanding relationships found in large textual datasets. Instead of forcing you to rely on keyword queries and to read long lists of documents, the Galaxies

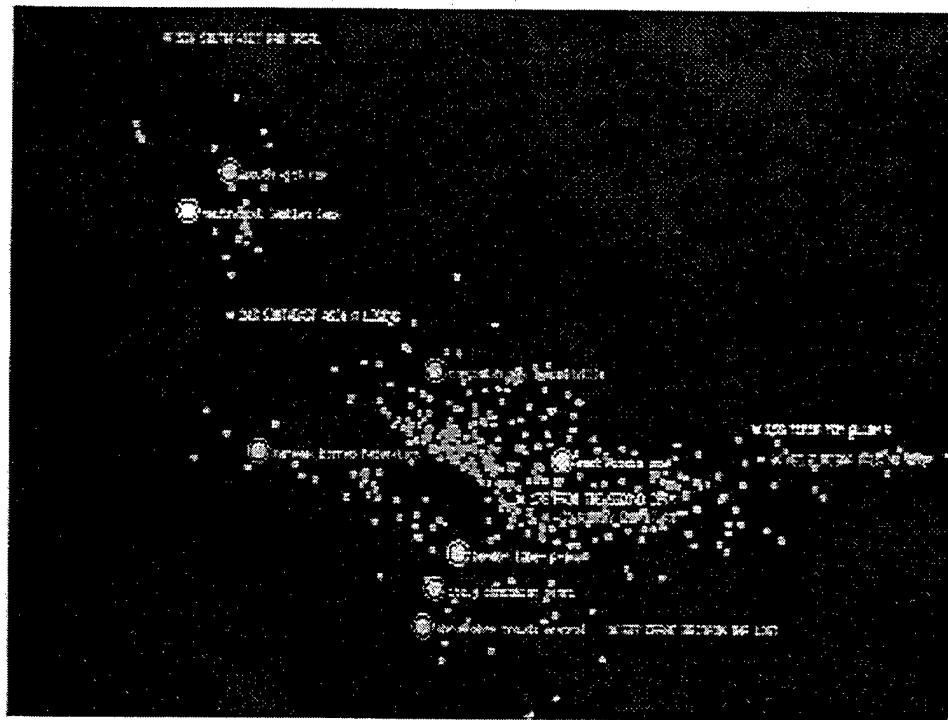


Figure 3.1 Galaxies Visualization

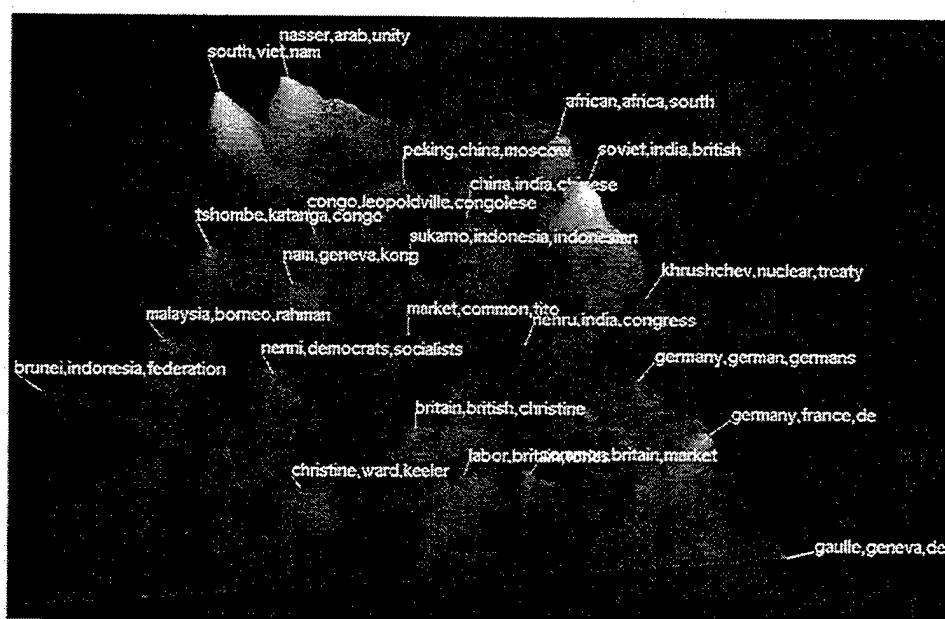


Figure 3.2 ThemeView Visualization

visualization gives you the ability to visually explore the relationships between all of your documents in a single visual metaphor--a picture of the entire textual dataset. In this visualization, you can visually locate and evaluate relationships between documents based on their proximity to each other. This allows you to quickly identify important documents in the dataset and to understand how all of the information is structured.

A ThemeView visualization is a 3-D feature map that shows where major concepts and themes are located in your dataset. Through a simple image, ThemeView visualizations can summarize the contents of an entire dataset, identify where major concepts are located, and detail the interrelationships between them. Combined with the thematic query tool, a ThemeView display becomes both a visual and interactive approach to exploring dataset contents. ThemeView visualizations are created from the Galaxies visualization by examining the contents of documents for words (called topics or themes) that best distinguish those documents from other regions of the dataset. When a localized region of documents containing similar themes occurs, a peak is assigned to that region. The height of the peak is dependent on how intensely the information is concentrated at that location. The net result of this approach is a graphic terrain in which strong themes are represented by high peaks. These peaks are easily identifiable and can provide insight into the information content of a dataset.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DATA

A. DESCRIPTION AND SELECTION

1. Naval Network Incidents

The Armed Services Computer Emergency Response Teams (CERTs) have established guidelines for categorizing network incidents by type. These categories vary in severity range from level one to seven (see Table 4.1). Level seven deals with virus activity and will not be addressed in this thesis. The indiscriminate nature of a computer

Category	Description
I	Unauthorized Root Access
II	Unauthorized User Access
III	Attempted Access
IV	Denial of Service
V	Poor Security Practice
VI	Probe or Scan
VII	Malicious Code

Table 4.1 Incident Category Description

virus, combined with the extreme difficulty in determining originating location and overall motivation for dispersal, would potentially allow for the introduction of outliers into the data set, which could skew the overall results.

NAVCIRT receives and collects Naval network incident data from IDS sensors and installation network administrators for all Navy networks. As stated, some of the data received is via IDS sensors, but the majority of incident data is gathered from emails and phone calls directly from installation network administrators to the NAVCIRT network analyst. Once received the incidents are investigated by the NAVCIRT analysts to determine the severity of the incident and its potential impact on the organization. If the incident is verified by the analyst it is categorized based on the JOINT CERT guidelines. Action will be taken by NAVCIRT personnel to correct or repair the affected system or area, and a notification report will be sent to JTFCND.

2. Foreign Broadcast Information Service

The Foreign Broadcast Information Service (FBIS) offers an extensive, in-depth collection of translations and transcriptions of open source information monitored worldwide on such diverse topics as military affairs, politics, the environment, societal issues, economics, and science and technology. The information is obtained monitoring radio, television, press, periodicals, books and other sources of unrestricted information such as databases and gray literature. These translations and transcriptions are known collectively as "FBIS Reporting." Information is collected from foreign media from all over the world. All collected data is translated and converted into a basic text electronic format. The data collected is available daily through the FBIS website, or in a quarterly update on CD ROM. The FBIS quarterly updates for the year 1999 will be used to provide the data source for the Visual Information Analysis tool SPIRE to analyze. The documents contained within the 1999-year data group will be from foreign media source across the globe.

B. DATA AGGREGATION

1. Naval Network Incidents

The Navy has been monitoring and recording network data since the mid-90's, however, prior to the year 1999 the data available was only from three Network Operating Centers. Data used in this thesis will be from validated incidents, categories I-VI, recorded during the year 1999 between January 1st and December 31st, because it is the first set of data to incorporate all four of the Navy's NOCs. During the entire year there were a total of 19,792 separate legitimate incidents. The data will be aggregated on a weekly basis throughout the year providing a total of 52 individual points of reference. Summary statistics (see Table 4.1) on the data indicate a maximum of 695 incidents and a minimum of 205 incidents per week throughout the year. There is a mean, or average value, of 380 with a standard deviation of 117. This would allow for the assumption that on the average there are approximately 380 incidents per week, give or take 117. But as you can see from Figure 4.1 this is slightly misleading as the number of incidents was lower in the first half of the year and higher in the second half.

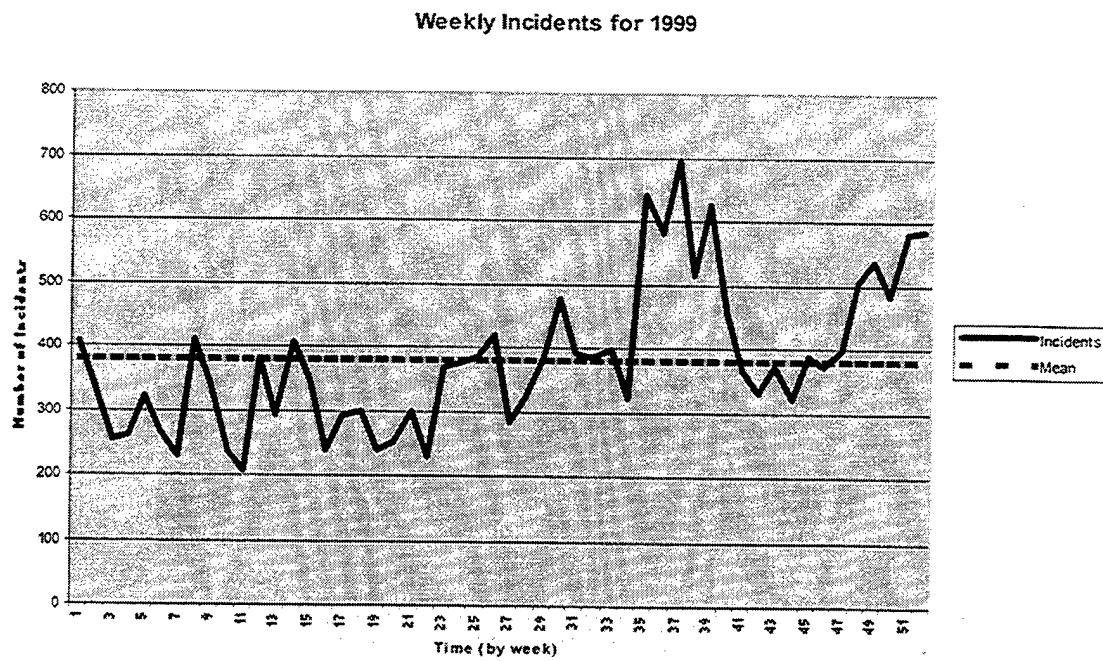


Figure 4.1 Network Incidents

Mean	380.6153846
Standard Error	16.19792163
Median	371.5
Mode	320
Standard Deviation	116.804874
Sample Variance	13643.37858
Kurtosis	0.238530438
Skewness	0.840289622
Range	490
Minimum	205
Maximum	695
Sum	19792
Count	52

Table 4.2 Summary Statistics

2. Media Visibility

The FBIS data is broken down into four quarters of approximately 3 months duration for the year 1999 (see Figure 4.2), beginning with December 30, 1998 and ending with December 31, 1999. Each quarter will be processed separately into its own dataset using SPIRE to create a Galaxies visualization from the datasets of raw, unformatted, unstructured text documents within the FBIS. The documents in a dataset are first processed by a text engine, which analyzes relationships among the words used and encodes those relationships numerically. In the Galaxies visualization, each document is represented as a point, called a *docustar*. Documents are logically grouped into relationships called *clusters*; these are represented in the Galaxies visualization as *cluster centroids*. Based on the numerical encoding computed by the text engine, documents and clusters of documents in the resulting Galaxies Visualization are positioned based on their relationship to every other document and cluster in the dataset. Proximity indicates relatedness, so documents and clusters that are similar in topical content will be shown with their docustars and centroids located closely together, while documents and clusters that are radically different in subject matter will have their symbols spaced farther apart in the visualization.

Each FBIS dataset is further delineated by separating the quarterly data into weekly increments using the Time Slicer feature (see Figure 4.2 FBIS Time Slices). The overall number of documents is indicated by the height of the bar representing each week within a particular quarter. The lower shaded portion of each bar is representative of the number of documents within the weekly totals that were indications of United States visibility.

Weekly Visibility for 1999

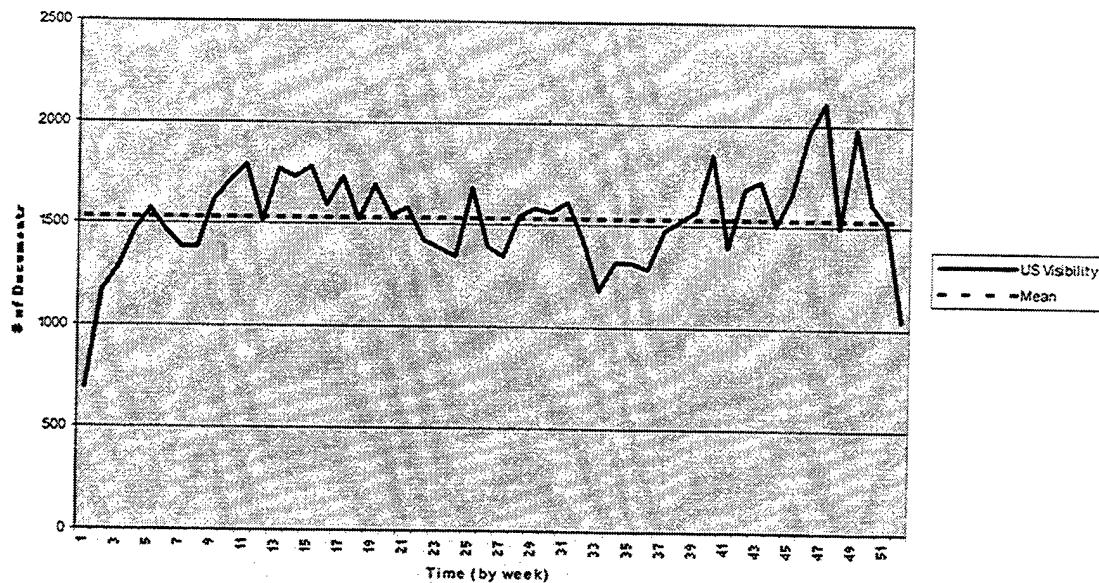
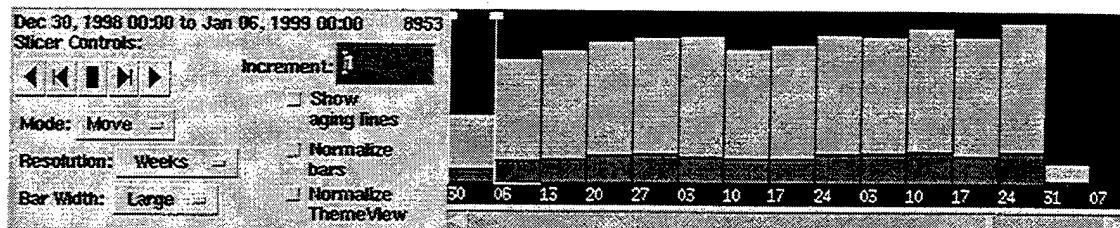


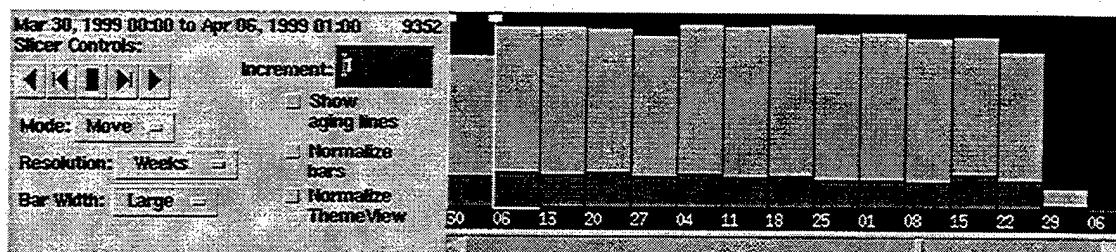
Figure 4.2 Document Visibility

Mean	1528.865385
Standard Error	33.35939436
Median	1533
Mode	1382
Standard Deviation	240.5580138
Sample Variance	57868.15799
Kurtosis	2.372130279
Skewness	-0.566510404
Range	1428
Minimum	684
Maximum	2112
Sum	79501
Count	52

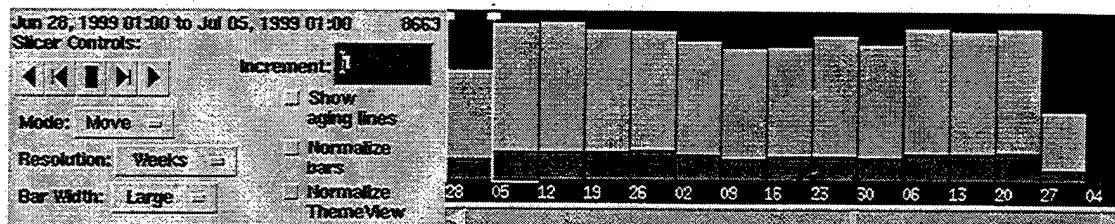
Table 4.3 Summary Statistics



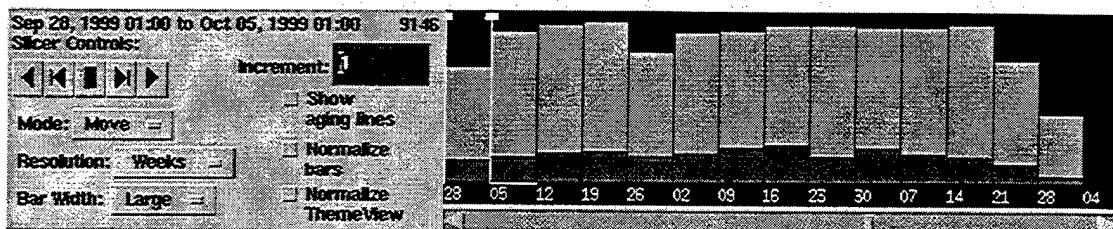
First Quarter 1999



Second Quarter 1999



Third Quarter 1999



Fourth Quarter 1999

Figure 4.2 FBIS Time Slices

THIS PAGE INTENTIONALLY LEFT BLANK

V. METHODOLOGY

A. OVERVIEW

The methodology for data collection in this study was a combination of literature review, personnel interviews and qualitative analysis of historical data. The data for the background chapter were collected through literature review of various reference materials. The data for the chapter on the Fleet Information Warfare Center and Pacific Northwest Laboratory were conducted through a literature review of internal documents and personnel interviews. The data for the analysis chapter was gathered based on a review of historical Navy network incident data and qualitative analysis of Foreign Broadcast Information Service reports based on documented foreign media from 1999 using the SPIRE VIA tool.

The theory of correlation was based on the time series concept that successive observations are usually not independent of one another. Examples occur in a variety of fields, ranging from stock prices or rainfall measured on successive days to population demographics measured annually. The purpose of analyzing a series of observations may be classified as description, explanation, and prediction. Description is the first step in any analysis, and will be the basis for this research. Plotting the data and obtaining basic statistics will make the identification of obvious features inconsistent with the data set as a whole, such as, trends, turning points and outliers, made visible (Chatfield, 1996). Explanation may generate a better understanding of the process that generated the series. Finally, prediction or forecasting (which are used interchangeably) is the process of

looking at past data and developing a systematic process for determining the most probable future outcome.

The relationship between two variables can be represented by the use of a mathematical formula. The Simple Linear Regression Model is one of the simplest and most easily understood formats.

$$Y_i = b_0 + b_1 \bullet X_i + \epsilon_i$$

Where

Y_i = independent variable

X_i = dependent variable

b_0 = Y intercept for the population

b_1 = slope for the population

ϵ_i = random error in Y for observation i

In this model, the slope (b_1) represents the amount of change in Y for a particular unit change in X and the Y intercept (b_0) represents the average value of Y when X equals 0.

B. HYPOTHESIS TESTING

Hypothesis testing is a phase of statistical analysis that utilizes a step-by-step process to obtain inferences concerning data, and determines the differences between observed results and the results that were expected based on some underlying theory. Testing begins with a theory or idea concerning a particular population of data. The initial hypothesis of this research is that there is a correlation between Naval Network

Incidents and US Visibility in the foreign media. Stated another way, the independent variable (visibility) affects the dependent variable (incidents) .

The first step is the development of the null hypothesis (H_0), which is always one of no difference. In this case, there is no effect by the independent variable on the dependent variable, or the slope coefficient (b_1) is equal to zero.

$$H_0: b_1 = 0$$

The slope represents the expected change in variable (y) per unit change in variable (x). The null hypothesis is always tested, always contains an equal sign and always refers to a specific value of the population parameter. If however the null hypothesis is proven false then something else must be true. This is planned for by the statement of the alternate hypothesis. The alternate hypothesis (H_1) is the opposite of the null hypothesis and represents the conclusion reached by rejecting the null hypothesis if there is sufficient evidence to determine it is untrue.

$$H_1: b_1 \neq 0$$

The process is designed to so that the rejection of the null hypothesis is based on evidence that the alternate hypothesis is far more likely. Even if the null hypothesis is true, sampling error will likely cause the statistic to differ from the parameter value. Hypothesis testing accounts for this by establishing of a clearly defined quantifying process, which determines the distribution for the statistic of interest and then computes a particular test statistic. Since sample distributions usually follow a normal distribution they can be used to determine the likelihood of the null hypothesis being true.

Distribution of the test statistic is divided into two regions, a region of rejection and a region of nonrejection. If the statistic falls within the region of nonrejection the null hypothesis cannot be rejected. The region of rejection on the other hand consists of values of the test statistic that are unlikely to occur if the null hypothesis is true. The size of the rejection region is determined by the critical value (which divides the nonreject region from the reject region) and the level of risk accepted. However, there is always the possibility when using a sample statistic that an incorrect conclusion will be reached. When conducting hypothesis testing there are two type of errors that can occur. A Type I error occurs if the null hypothesis is rejected when it is actually true. The probability of committing a Type I (α) error is referred to as the level of significance. The level of significance is determined before testing is performed and is usually at a level of .05 or smaller. Determination of the level of significance will establish the size of the rejection region. A Type II (β) error occurs if the null hypothesis is not rejected and it is actually false. The risk of a Type II error is determined by the difference between the hypothesized and actual values of the parameter.

C. AUTOCORRELATION

Statistical tests depend on the randomness of the data being analyzed. This randomness constitutes one of the four assumptions that underlie all measurement processes. If the data is not random then the validity of the statistical conclusions become suspect. When looking at time series, if successive error terms are independent of one another then so are successive observations. However, if the error terms are statistically dependent then so are the successive observations. The observations would be called autocorrelated. The assumption of independency of errors is often violated

when data is collected over a period of time because a residual at one point in time may tend to be similar to residuals at adjacent points in time. Measuring autocorrelation is a technique that checks for randomness in a data set and can determine whether an observation is related to an adjacent observation. The Durbin-Watson (D) statistic can detect and measure the correlation between each residual and the residual for the time period immediately preceding the one of interest. If the residuals are not correlated, the value of D will be close to 2. A value of D close to 0 denotes positive autocorrelation, while a value of D close to 4 indicates negative autocorrelation. When attempting to use the Durbin-Watson technique the main issue is when does the level of autocorrelation fall sufficiently below 2 to cause concern about the validity of the data. This determination is dependent on the number of observations (n), the number of independent variables (p) in the model and the level of significance (α). Based on these three items a standard table for the Durbin-Watson Statistic will provide two critical values. The value (d_L) represents the lower critical value. If D is below d_L then there is evidence of positive autocorrelation. The value (d_U) represents the upper critical value. If D is above d_U then there is no evidence of autocorrelation among the residuals. If the value of D is between d_L and d_U then it is impossible to make a definite conclusion about the data.

In the event that a data set is determined to be autocorrelated it is necessary to account for the correlation between residuals by transforming the data prior to conducting any regression analysis. This transformation is accomplished by using the previously determined Durbin-Watson statistic to find the estimate of the autocorrelation parameter (r) and transform both the independent (y) and dependent (x) variable.

Estimate of the autocorrelation parameter:

$$r = 1 - (D/2)$$

Transformation of independent variable (y to y*):

$$y_1^* = (\sqrt{1 - r^2}) \bullet y_1$$

$$y_2^* = y_2 - (r \bullet y_1)$$

$$y_3^* = y_3 - (r \bullet y_2)$$

↓

$$y_n^* = y_n - (r \bullet y_{n-1})$$

Transformation of the dependent variable (x to x*):

$$x_1^* = (\sqrt{1 - r^2}) \bullet x_1$$

$$x_2^* = x_2 - (r \bullet x_1)$$

$$x_3^* = x_3 - (r \bullet x_2)$$

↓

$$x_n^* = x_n - (r \bullet x_{n-1})$$

D. CORRELATION

Correlation analysis measures the relationship between two items. You can use correlation analysis in two basic ways: to determine the predictive ability of an indicator and to determine the association between two items. When comparing the correlation between two items, one item is called the "dependent" variable and the other the

"independent" variable. The goal is to see if a change in the independent will result in a change in the dependent. This information helps you understand an indicator's predictive abilities. When comparing the correlation between a network incident and media visibility, a high positive coefficient (e.g., +0.70) tells you that a change in the visibility will usually predict a change in the level of incident. A high negative correlation (e.g., -0.70) tells you that when the visibility changes, the incident level will usually move in the opposite direction. The coefficient of correlation (r) is obtained from the coefficient of determination (r^2).

$$r = \sqrt{r^2}$$

The correlation coefficient can range between ± 1.0 (plus or minus one). A coefficient of +1.0, a "perfect positive correlation," means that changes in the independent item will always result in a change in the dependent item. A coefficient of -1.0, a "perfect negative correlation," means that changes in the independent item will always result in a change in the dependent item, but the change will be in the opposite direction. A coefficient of zero means there is no relationship between the two items and that a change in the independent item will have no effect in the dependent item. A low correlation coefficient (e.g., less than ± 0.10) suggests that the relationship between two items is weak or nonexistent. A high correlation coefficient (i.e., closer to plus or minus one) indicates that the dependent variable will usually change when the independent variable changes. The direction of the dependent variable's change depends on the sign of the coefficient. If the coefficient is a positive number, then the dependent variable will move in the same

direction as the independent variable; if the coefficient is negative, then the dependent variable will move in the opposite direction of the independent variable.

The coefficient of determination measures the proportion of variation in the dependent variable (network incidents) that is explained by the independent variable (US visibility).

$$r^2 = \text{regression sum of squares (SSR)} \div \text{total sum of squares (SST)}$$

The coefficient of determination is a measurement, usually represented by a percentage value that indicates either a high or low linear relationship between the two variables based on the ability of the regression model to account for changes in the independent variable. A higher coefficient of determination (r^2) would be statistically significant in indicating a strong linear relationship between the variables and would be statistically equivalent to the slope (b_1) being significantly different from zero.

$$r^2 > 0 \approx b_1 > 0$$

To fully understand how the independent variable predicts the dependent variable it is necessary to determine the total sum of squares (SST). SST is a measure of variation on the independent variable values around their mean and is equal to the sum of the squared differences between each observed value (Y_i) and the mean value of (\bar{Y}).

$$SST = \sum (Y_i - \bar{Y})^2$$

And the regression sum of squares (SSR). SSR is attributed to the relationship between the independent and dependent variable, and is equal to the sum of the squared differences between each predicted value of (\hat{Y}_i) and the mean of (\hat{Y}).

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2$$

E. PROCESS

The first step is to aggregate the Naval network data based on the desired time, in this case a weekly aggregate, then determine the summary statistics and produce a timeline graph. Next, convert the FBIS foreign media data into a Galaxy visualization format using the SPIRE visual information analysis software. Conduct a visual review of the galaxy display to identify any obvious or unusual patterns. Third, utilize SPIRE's analysis tools to better understand the relationships represented in the Galaxies visualization. Specifically, the analysis of data will be as follows; 1) Use the *Query* tool, with the key phrase "United States" to identify all documents that are related to or reference the United States. 2) Use the *Group* tool to create a sub-set of documents within the FBIS data. 3) Use the *Time Slicer* tool to separate the United States grouped documents for the entire year by week of occurrence. Finally, combine the network incident data with the FBIS data to produce a timeline graph for 1999. Conduct visual analysis of the plotted incident data and FBIS data to determine any obvious features or patterns. Specifically, any group of FBIS visibility data that immediately precedes an increase or rise in the level of network incidents. Alone, an individual document or group of FBIS visibility documents that precede a rise in network incidents are probably unrelated. However, if this pattern repeats itself over time then that may indicate a potential connection or correlation. Develop a null and alternate hypothesis to support the theory of correlation. Conduct the Durbin-Watson statistic using the PH-Stat statistical add-in package for Microsoft Excel to determine if the data set is autocorrelated. If autocorrelated, utilize the estimate of the autocorrelation parameter to

transform the independent and dependent variable to account for residual correlation. Reconduct the Durbin-Watson statistic to confirm that autocorrelation has been eliminated. Finally, utilize Microsoft Excel to conduct regression analysis and determine the coefficient of correlation (r) between the two data sets.

F. SPIRE ANALYSIS TOOLS

The analysis tools allow you to perform queries, group documents, view slices of your dataset based on document timestamps and on Boolean query logic, and retrieve statistically selected keywords called *gisting terms* for different regions of the dataset. By using these tools, you can explore relationships in your dataset that may not be immediately obvious in the Galaxies visualization.

The Query Tool is designed to provide you with the ability to query a dataset using three different techniques; *Vocabulary Word Queries*: applying Boolean logic to vocabulary words, *Exact Phrase Queries*: searching with controlled scope and case sensitivity and *Queries by Example*: measuring proximity of words in high-dimensional space. Using this tool is a good way to begin to locate specific information in a dataset. By entering a query, you can quickly find concentrations of documents containing topical areas of interest, and identify related documents that are located nearby but may not contain the query terms or phrases. The Query tool also enables you to automatically select and group the documents located by a query. Grouping tool is a convenient way of organizing sets of documents in the Galaxies visualization. By assigning documents to a group, you can rapidly retrieve query results, use set operations to examine the union, difference, and intersection of document sets, and reselect sets of groups. With the Time Slicer you can explore how your dataset evolves over time. The Time Slicer works by

examining the timestamp in each document, and partitioning the dataset into discrete time slice views. Document counts for each time slice are depicted in a histogram. By interacting with this histogram, you can rapidly "slice" through a dataset and watch your information evolve over time. The Time Slicer can partition a dataset only if timestamps are available when the dataset is initially processed, with the granularity of the slices dependent on the granularity of the timestamps. The Time Slicer is designed to slice time intervals ranging from years down to individual minutes.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. ANALYSIS

Four charts were developed to indicate weekly FBIS document distribution for each quarter of 1999. The charts (Figures 6.1 through 6.4) display the total number of documents in a given week, as compared to the number of documents that indicate United States visibility within the same week. The first step in analysis of a time series is to plot the data to identify trends. A chart (see Figure 6.5) was constructed to display the level of network incidents and United States visibility within those documents during the year 1999. An initial visual inspection of the Incident and Visibility graph provides no clear or easily identifiable pattern. The only obvious similarity is that both sets of data have an increasing mean value, meaning that the number of events for both sets of data continues to grow in size throughout the year. As the data was collected quarterly and displayed in a chronological fashion for the year, the behavior of the data sets might be due to some seasonality influence. However, there is no information or data to support this theory.

Based on the theory that there is a correlation between network incidents and US visibility in the foreign media, a hypothesis statement was formed. The null hypothesis is that there is no impact by the independent variable visibility (y) on the dependent variable incidents (x). The slope (b_1), which is the change in (y) per unit change in (x), is zero. The alternate hypothesis will be the converse.

$$H_0: b_1 = 0$$

$$H_1: b_1 \neq 0$$

The level of significance (α) established for this hypothesis will be .01. This will reduce the probability of committing a Type I error (rejecting the null when it is true), while accepting a greater amount of risk of committing a Type II error (not rejecting the null when it is false).

Prior to any regression analysis the Durbin-Watson test was run on the data set to determine the D value. This D value was then compared against critical values in a Durbin-Watson Critical Values table (Table E.8, Appendix E-17, ref: #3) to see if any autocorrelation between variable residuals was present. The Durbin-Watson table provides the upper and lower critical values based on the following parameters, the level of significance (α), the sample size (n), and the number of independent variables (p).

$$\alpha = .01, n = 52, p = 1$$

$$d_L = 1.32$$

$$d_U = 1.40$$

The Prentice-Hall Statistical (PH Stat) add-in program for Microsoft Excel analyzed the initial variables and produced a D value of .65.

$$D = .65 < 1.32$$

Because the D value is less than the lower critical value it can be concluded that positive autocorrelation exists among the residuals. To eliminate the autocorrelation it was necessary to transform the initial observations. This transformation was conducted by determining the estimate of the autocorrelation parameter (r) and utilizing this value to adjust both the independent and dependent variables.

$$1 - (D / 2) = r$$

$$1 - (.65 / 2) = r$$

$$1 - .32 = r$$

$$r = .68$$

Once the observations had been transformed and autocorrelation accounted for, the Durbin-Watson test was run again on the adjusted data using the same initial parameters to validate the transformation.

$$\alpha = .01, n = 52, p = 1$$

$$d_L = 1.32$$

$$d_U = 1.40$$

This time the Durbin-Watson test produced a D value of 2.26. When the D value is greater than 2, indicating negative autocorrelation, the Durbin-Watson upper and lower critical values must be adjusted by subtracting d_U and d_L from 4.

$$4 - d_L = d_L^* \quad 4 - d_U = d_U^*$$

$$4 - 1.32 = 2.68 \quad 4 - 1.40 = 2.60$$

$$D = 2.26 < 2.60$$

Because the D value is lower than four less the upper critical value from the table it can be concluded that the negative autocorrelation that is present is not statistically significant, and that regression analysis and its estimated parameters will not be biased.

Regression analysis of the transformed data provided a slope coefficient of -.06.

$$b_1 = -.06$$

The level of significance (α) of .01 established a confidence level of .99 or 99% for this hypothesis. Based on a 99% confidence level the upper and lower critical values were determined.

$$\text{lower 99\%} = -.24 \quad \text{upper 99\%} = .11$$

$$-.24 < -.06 < .11$$

This established a region of nonrejection between -.24 and .11. Since -.06 falls within the region of nonrejection the null hypothesis is not rejected.

Using Microsoft Excel's correlation analysis tool, the coefficient of correlation (r) was determined to be (.12) approximately.

$$r = \sqrt{r^2}$$

$$r^2 = 41812 \div 2951276 = .014$$

$$r = \sqrt{.014} = .118$$

$$.118 \approx .12$$

As stated, correlation attempts to establish the strength of association between numerical values. In this case, a result of (.12) indicates an extremely low correlation. The proximity of the coefficient to the value zero indicates that there is almost a non-existent correlation between the level of network incidents and US visibility in the foreign media.

This conclusion is confirmed by the confidence interval on b_1 .

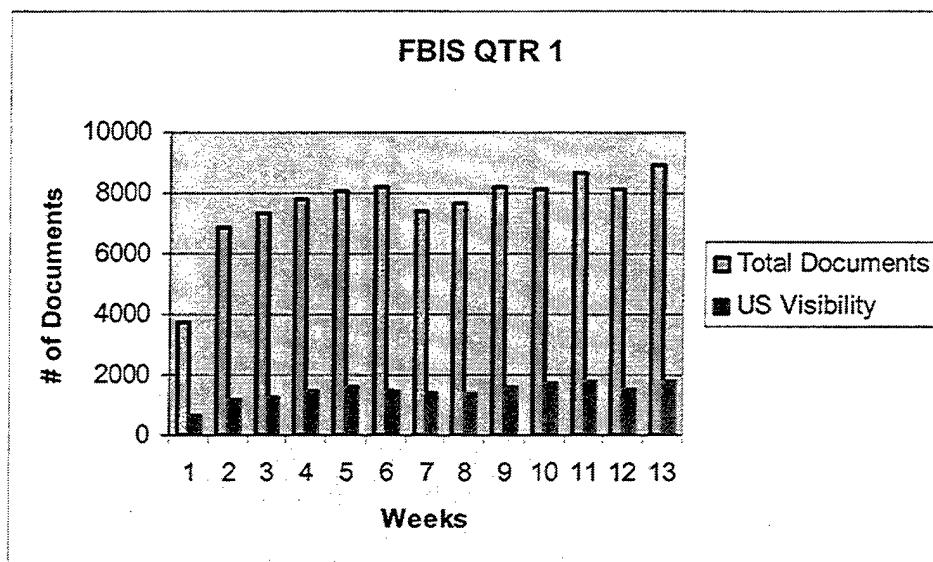


Figure 6.1 FBIS Documents by Quarter

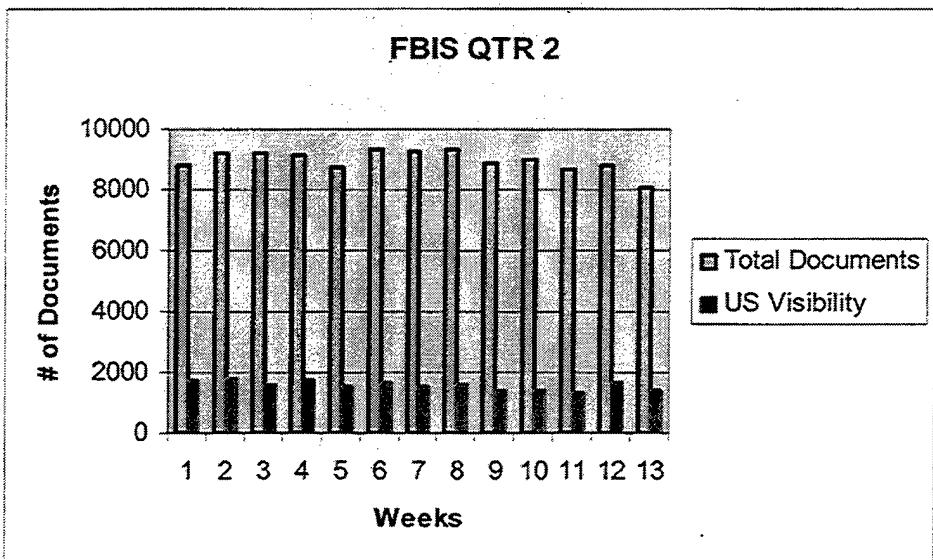


Figure 6.2 FBIS Documents by Quarter

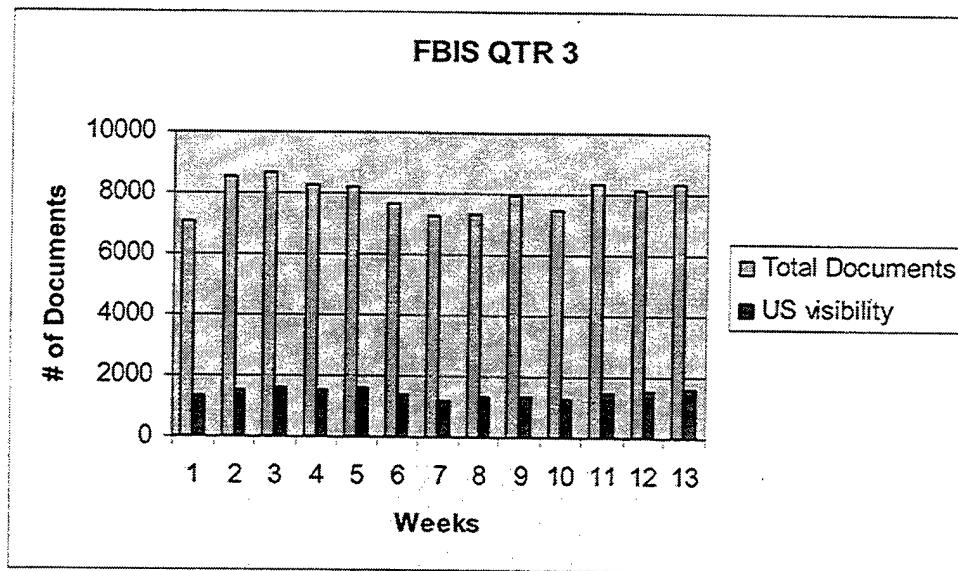


Figure 6.3 FBIS Documents by Quarter

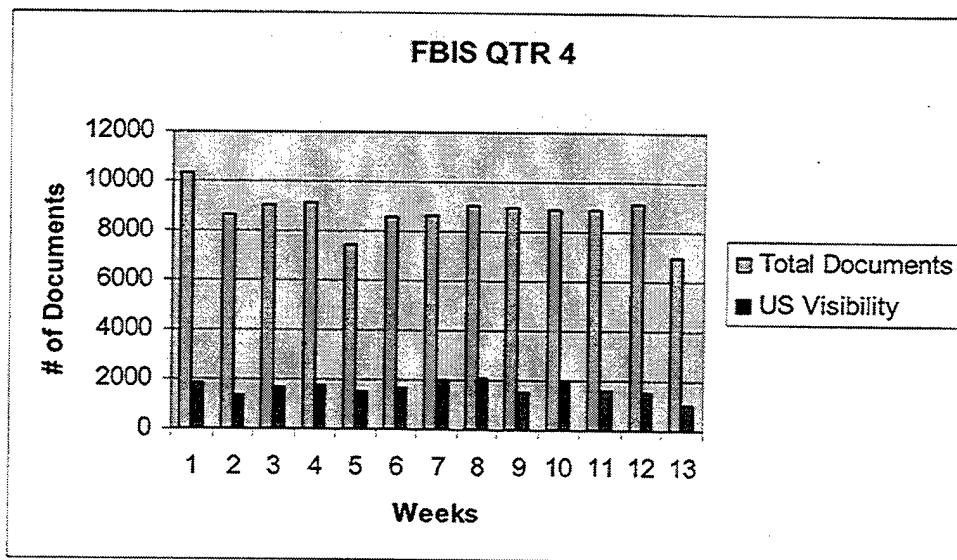


Figure 6.4 FBIS Documents by Quarter

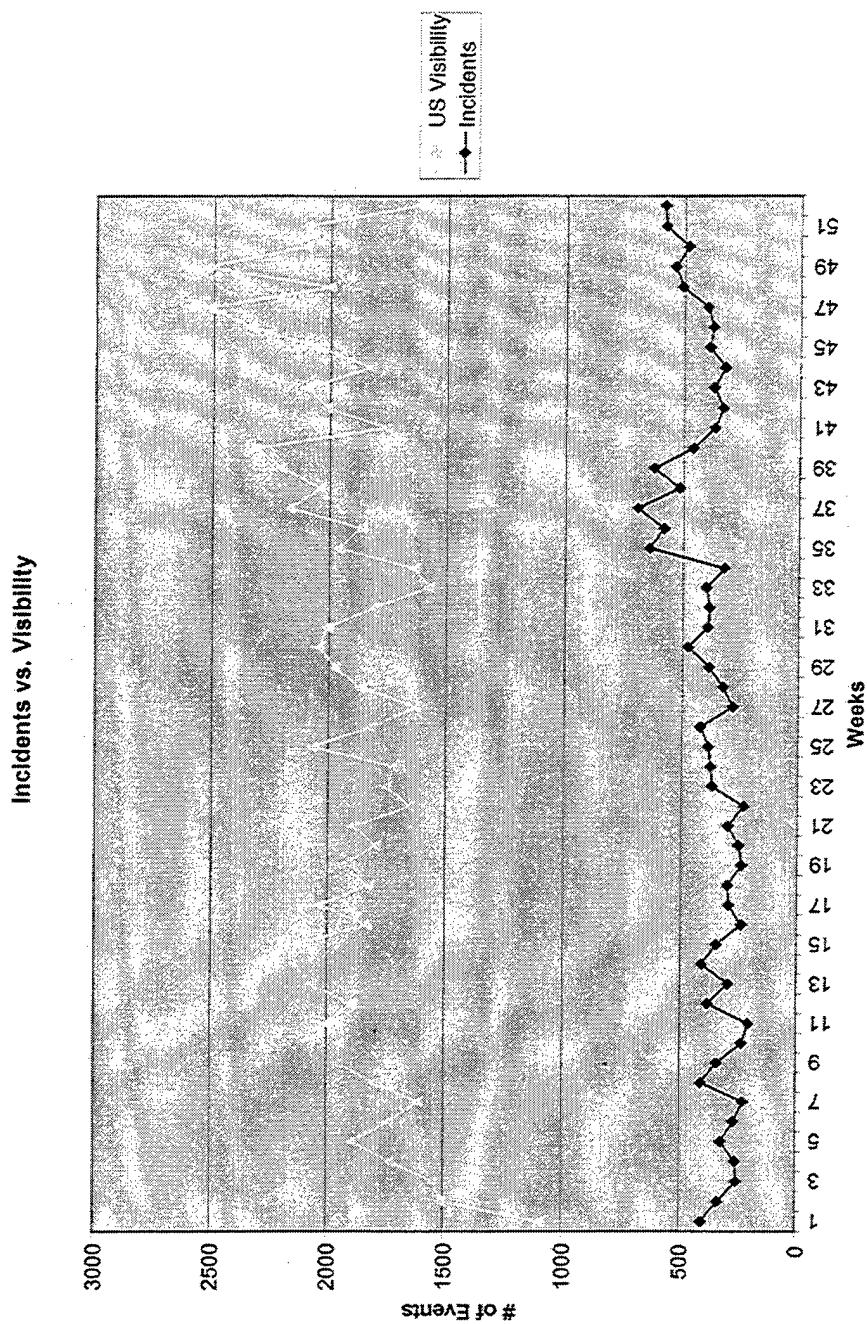


Figure 6.5 Incident & Visibility Graph

THIS PAGE INTENTIONALLY LEFT BLANK

VII. CONCLUSIONS AND RECOMMENDATIONS

A. SUMMARY

Intrusion detection is a process that is reactive in nature. The Navy currently faces the growing challenge of providing Information Assurance in a constantly changing technical environment. Under these conditions, NAVCIRT requires tools to quantify the impact of external factors on their communication network, identify potential negative impacts in a more timely and efficient manner and develop a process to neutralize them.

This thesis examined Naval network incident data and Foreign Broadcast Information Service data from 1999, and attempted to do the following:

1. Determine if a methodology could be developed to analyze incident and media data.
2. Determine if a correlation exists between network incidents on Navy networks and visibility of the United States in the foreign media.

A time series approach was utilized to analyze an aggregate of network incidents throughout the year, and compare them to United States visibility within the foreign media. The foreign media events were selected based on the amount of articles during the year 1999 that provided "visibility" by directly referencing the United States. Both sets of data were aggregated on a weekly basis. Analysis of the Naval network incident data and Foreign Broadcast Information Service media data for the year 1999 provided no indication that a correlation existed. The correlation coefficient between the two sets of data was approximately .12 suggesting that the relationship between the two is weak or

non-existent. In addition, the hypothesis that the slope coefficient was equal to zero was proven true showing that a change in the independent variable had no effect on the dependent variable.

B. CONCLUSIONS

The extremely low correlation coefficient between Naval network incidents and visibility within the foreign media supports the null hypothesis that a change in the independent variable visibility has no effect on the dependent variable incidents. Clearly, the level of Naval network incidents continued to increase throughout the year, but this can in no way be attributed to the amount of United States visibility within the foreign media at any particular time. Though the analysis represented in this thesis may represent a descriptive tool for what occurred during the period studied, it lacks the forecasting accuracy to provide legitimate opportunity for correlation analysis. However, it was possible to develop a process for analyzing network incidents and media visibility. And although this research was unable to provide any indication of correlation between the chosen data sets, the methodology itself offers potential for further refinement and application, and may prove to be a useful tool in providing for a more proactive process of network defense in the future.

C. RECOMMENDATIONS FOR FURTHER RESEARCH

Recommendations for improvements to correlation research fall in three categories: data collection, timeline analysis and methodology refinement. Future studies should attempt to utilize incident data that has a specific originating source. In the investigation of network incidents it is possible to determine not only the country of origin, but in many cases the responsible individual. Information of this nature is beyond

the scope of this research and in many cases is sensitive or classified in nature. In addition, limiting the visibility data to media content that originates from the same country or geographical region as the incident data will provide for a more sound and logical attempt at correlation. Another technique at refinement would be to focus all data down to daily observations, which would allow for a more detailed and in-depth analysis. As more and more data becomes available the timeline may also be extended to include more than just one year in the sample size. A larger population would provide for a more complete and accurate statistical analysis. In addition, while the SPIRE visual information analysis tool was vital in mining the FBIS data for the general numbers, the actual interface to extract the specific events was a completely manual process. It would not only speed the research process to automate the data mining, data extraction and data analysis process, but provide for the ability to analyze larger or even multiple sets of data. It may be possible to completely automate the process of FBIS data mining and correlation analysis through the development of specialized software.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

1. Manheim, Jarol B., *Empirical Political Analysis: Research Methods in Political Science*, Longman, 1991.
2. Chatfield, Chris, *The Analysis of Time Series: An Introduction*, Chapman and Hall, 1996.
3. Levine, David M., *Statistics for Managers: using Microsoft Excel*, Prentice Hall, 1999.
4. Bowerman, Bruce L., *Time Series and Forecasting: An Applied Approach*, Duxbury Press, 1979.
5. Burger, Eric C., *A Multivariate Time Series Analysis of U.S. Army Recruiting*, Naval Postgraduate School, 2000.
6. SPAWAR, *Users's Logistic Support Summary (Draft)*, 1999.
7. Joint Pub 3-13, *Joint Doctrine for Information Operations*, 1998.
8. JV 2020, *Joint Vision 2020*, US Government Printing Office, 2000.
9. OPNAV Instruction 2201.2
10. www.chinfo.navy.mil.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
8725 John J. Kingman Road, Suite 0944
Fort Belvoir, VA 22060-6218
2. Dudley Knox Library 2
Naval Postgraduate School
411 Dyer Road
Monterey, CA 93943-5101
3. Fleet Information Warfare Center 2
2555 Amphibious Drive, Bldg 1265
Norfolk, VA 23521-3225
4. Naval Information Warfare Activity 2
9800 Savage Road
Fort Meade, MD 20755-6000
5. Joint Information Operations Center 2
Z Hall Blvd, Ste 217
San Antonio, Texas 78243-7008
6. LT Ray Buettner 8
Code IW/BR
Naval Postgraduate School
Monterey, CA 93943-5101